A general criterion for image similarity detection

Frédéric Cao, Patrick Bouthemy

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A general criterion for image similarity detection

Frédéric Cao and Patrick Bouthemy

N°5620

Juillet 2005

Systèmes cognitifs

ISSN 0249-6399 ISRN INRIA/RR--5620--FR+ENG

A general criterion for image similarity detection

Frédéric Cao * and Patrick Bouthemy †

Systèmes cognitifs
Projets Vista

Rapport de recherche n°5620 — Juillet 2005 — 16 pages

Abstract: A novel and general criterion for image similarity is introduced, based on the comparison of grey level gradient direction at randomly sampled points. It is mathematically proved that it is possible to compute a fully automatic and robust threshold to detect that two images have a common cause, which can be taken as a definition of similarity. Analytical estimates of the necessary and sufficient number of sample points are also given. Similar pairs of images are detected *a contrario*, by rejecting an hypothesis that resemblance is due to randomness, which is far more easy to model than a realistic degradation process. The method proves very robust to noise, transparency and occlusion.

Key-words: Image comparison, *a contrario* detection, number of false alarms

(Résumé : tsvp)

* Frederic.Cao@irisa.fr
† Patrick.Bouthemy@irisa.fr

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : 02 99 84 71 00 - International : +33 2 99 84 71 00
Télécopie : 02 99 84 71 71 - International : +33 2 99 84 71 71

Un critère général de similarité entre images

Résumé : Cette note décrit un critère général de comparaison d'images, basé sur la comparaison des directions du gradient en niveau de gris, échantillonnées aléatoirement dans l'image. On montre qu'il est possible de calculer un seuil permettant de détecter que la ressemblance entre deux images ne peut pas être due au hasard, ce qui peut être pris comme définition de la similarité. Il est également possible d'estimer le nombre de points nécessaire et suffisant permettant de passer le test. Des images semblables sont donc détectées *a contrario*, ce qui est bien plus simple que de modéliser un processus de dégradation réaliste. Des expériences montrent que les seuils sont très robustes au bruit, aux occlusions ou à la transparence.

Mots-clé : Comparaison d'image, détection *a contrario*, nombre de fausses alarmes

1 Introduction

Establishing that two images (or parts of images) are similar is a general concern in image analysis and computer vision. It is involved in a number of problems or applications [1] such as image matching, displacement computation, stereovision, change detection, image or video retrieval... In this paper, we answer the following question : with which degree of certainty can we assert that two images are similar? A second question is: can we compute "universal" thresholds to decide to match two images? This problem is very difficult in full generality since image similarity should be define up to a large group of invariance: geometric deformation, contrast change, scaling, occlusion, transparency, noise, etc... In this work, we do not seek the largest group of invariance (we shall give hints how to achieve this), but rather concentrate on the second step. More precisely, given two images or pieces of images that are supposedly registered, how to come to a very robust measure of similarity. Even on whole images, this problem has several applications: image retrieval that consists to check whether or not an image is present in a video stream or in a database, verification of the accuracy of registration. The proposed solution is extremely stable with respect to noise (it still works with an additive Gaussian noise with standard deviation 30 or a 50% impulse noise). It is based on statistical arguments exploiting very simple information computed on the image intensities. The search is totally processed online and is very efficient (10 frames/s on a 2.4GHz PC, with no optimization). Since it only relies on the direction of the image gradient, the method is contrast invariant.

2 Related work

The statistical arguments we introduce can be related to the work of Lisani and Morel [8]. Their approach uses the direction of the gradient of a grey level image, and they detect local changes in registered stereo pairs of satellite images. Our method is dual since, on the contrary, we use the gradient direction in both images to decide that they have much spatial information in common. Detection thresholds are computed by using an *a contrario framework*, as introduced by Desolneux, Moisan and Morel [2]. More ancient work [15] used the same kind of ideas but detection thresholds were not computed. Other image features widely used are SIFT descriptors [10, 9] which are basically local direction distributions. Nevertheless, the indexing and comparison of descriptors is achieved by a nearest-neighbor procedure. Hence, there are no automatic decision thresholds, which is precisely our main concern. On the other hand, we think that our methodology can be adapted to the comparison of SIFT features. Basically, our method consists in sampling random points in two images and counting the number of points such that the difference of the gradient direction

is less than a given threshold. If this number is large enough, then images have certainly a common cause. Remark that contrarily to methods as RANSAC [4], we do not try to estimate any registration parameters, because probabilities will be computed in a model representing the absence of similarity (background model, in the statistical meaning). Some similar idea can be found in [5] where the authors study the influence of "conspiracy of random".
The paper is organized as follows. In Sect. 3, we define a criterion yielding a recognition threshold of an image in a database. Even though we use an hypothesis testing formalism, we show that decision only relies on the likelihood of one hypothesis (which is that the two compared images are not the same). The test compares the image gradient direction at some random points. In Sect. 4, we show that this number of sample points can be chosen to maintain a probability of detection very close to 1, when we assume white Gaussian noise. However, we insist that detection does not rely on such an assumption. It will be observed that, in practice, the required number of samples is seldom above a few hundreds, even for quite important noise. In Sect. 5, we give numerical applications for typical images, and Gaussian or impulse noise. We discuss numerical artifacts, as gradient quantization, and report some experiments on image retrieval in databases of typically $10,000$ images. It will be shown that the method is robust with respect to transparency and occlusion.

3 A contrast invariant image comparison method

In what follows, we always assume that images are grey level valued with size $N \times N$. Let u and v be two images. For any point x, let us denote by $\theta_u(x)$ and $\theta_v(x)$ the directions of the gradient of u and v at point x. Let us denote by $D(x) = d_{\mathbb{S}^1}(\theta_u(x), \theta_v(x))$ the geodesic distance between $\theta_u(x)$ and $\theta_v(x)$ on the unit circle $\mathbb{S}^1$. It is a real value in $[0, \pi]$. Since we want this measure to be accurate, we only consider points where both image gradients are large enough (larger than 5 in practice). Now, two images differing from a contrast change have the same gradient direction everywhere, and this is what is detected in the following.

Even though the proposed method is not a classical hypothesis testing, let us formulate it this way. From the observations of the values of $D(x)$, we want to select one of the two following hypotheses: $\mathcal{H}_0$: u and v are the same image up to some degradation; $\mathcal{H}_1$: u and v are different images. Modeling Hypothesis $\mathcal{H}_0$ is equivalent to model the type of degradation that can lead from u to v, and only very simplistic models are usually at hand. In Section 4, we will make such an assumption to discuss the detection rate. In an image retrieval application, v can belong to a database of typically 10^6 images (10 hours of video). Hence, false alarms (that is accept H_0 while H_1 actually holds) have to be controlled, else the system will become unpractical. Because of the large size of the database, this implies

that it is necessary to compute very small probabilities of false alarms. The proposed method is to base the decision only on $\mathcal{H}_1$, which is far more easy to model. It allows us to attain very small probabilities of false alarm. Moreover, there is no need to compare the likelihood of the two hypotheses, since we can derive automatic thresholds on the likelihood of $\mathcal{H}_1$, which allows us to reject it very surely.

Hypothesis $\mathcal{H}_1$ models the absence of similarity. Thus, the following assumption is made: the families of $D(x)_{x\in[1,N]^2}$ are independent, identically distributed in $[0,\pi]$. This probabilistic model will be called the *a contrario* model or background model. The principle of the detection is to compute the probability that the real observation has been generated by the *a contrario* model. When this probability is too small, the independence assumption of the two images is rejected and similarity is detected a contrario.

Let us fix $\alpha \in (0,\pi)$, and let $q_\alpha = \frac{\alpha}{\pi}$. For any set of distinct points $\{x_1,\dots,x_M\}$, the probability, under $\mathcal{H}_1$, that at least k among the M values $\{D(x_1),\dots D(x_M)\}$ are less than α is given by the tail of the binomial law

$$B(M,k,q_\alpha) = \sum_{j=k}^{M} \binom{M}{j} q_\alpha^j (1-q_\alpha)^{M-j}.$$

Definition 1 *Let $0 \leqslant \alpha_1 \leqslant \dots \leqslant \alpha_L \leqslant \pi$ be L values in $[0,\pi]$. Let u a real valued image, and $x_1,\dots x_M$, M distinct points. Let us also consider a database $\mathcal{B}$ of $N_\mathcal{B}$ images. For any $v \in \mathcal{B}$, we call number of false alarms of (u,v) the quantity*

$$NFA(u,v) = N_\mathcal{B} \cdot L \cdot \min_{1\leqslant i\leqslant L} B(M,k_i,q_{\alpha_i}), \tag{1}$$

where k_i is the cardinality of

$$\{j,\ 1 \leqslant j \leqslant M, d(\theta_u(x_j),\theta_v(x_j)) \leqslant \alpha_i\}.$$

We say that (u,v) is ε-meaningful, or that u and v are ε-similar if $NFA(u,v) < \varepsilon$.

The interpretation of this definition will be made clear after stating the following proposition.

Proposition 1 *For a database of $N_\mathcal{B}$ images such that the gradient direction difference with a query u has generated by the background model, the expected number of v such that (u,v) is ε-meaningful is less than ε.*

Proof. For any v, (u,v) is ε-meaningful, if there is at least $1 \leqslant i \leqslant L$ such that $N_\mathcal{B} \cdot L \cdot B(M,k_i,q_{\alpha_i}) < \varepsilon$. Let us denote by $E(v,i)$ this event. By definition, its probability $P(E(v,i)) \leqslant \frac{\varepsilon}{L\cdot N_\mathcal{B}}$. The event $E(v)$ defined by "(u,v) is ε-meaningful" is $E(v) =$

$\cup_{1 \leqslant i \leqslant L} E(v,i)$. Let us denote by $\mathbb{E}_{\mathcal{H}_1}$ the mathematical expectation under the background model. Then

$$\begin{aligned}
\mathbb{E}_{\mathcal{H}_1}\left(\sum_{v \in \mathcal{B}} \mathbf{1}_{E(v)}\right) &= \sum_{v \in \mathcal{B}} \mathbb{E}_{\mathcal{H}_1}(\mathbf{1}_{E(v)}) \\
&\leqslant \sum_{\substack{v \in \mathcal{B} \\ 1 \leqslant i \leqslant L}} P_{\mathcal{H}_1}(E(v,i)) \\
&< \sum_{\substack{v \in \mathcal{B} \\ 1 \leqslant i \leqslant L}} \frac{\varepsilon}{L \cdot N_{\mathcal{B}}} = \varepsilon. \qquad \square
\end{aligned}$$

Thus, Def. 1 together with Prop. 1 mean that there is in average less than ε images v in the database $\mathcal{B}$ that match with u by chance, that is to say, when $\mathcal{H}_1$ holds. Under this hypothesis, any detection must be considered as a false alarm (hence the denomination). Thus, it is chosen to eliminate any observation having a frequency of the order of ε in the *a contrario model*. In Sect. 5.1, it will be checked that Hypothesis $\mathcal{H}_1$ is sound for two unrelated images.

Even though this is theoretically simple, it may be difficult to numerically evaluate the tail of the binomial law. A sufficient and more tractable condition of meaningfulness is given by the following classical result, first proved by Hoeffding [6].

Proposition 2 *Let $H(r,p) = r \ln \frac{r}{p} + (1-r) \ln \frac{1-r}{1-p}$, be the relative entropy of two Bernoulli laws with parameters r and p. Then, for $k \geqslant Mp$,*

$$B(M,k,p) \leqslant \exp\left(-M \cdot H\left(\frac{k}{M}, p\right)\right). \tag{2}$$

This inequality leads to the following sufficient condition of meaningfulness.

Corollary 1 *If*

$$\max_{\substack{1 \leqslant i \leqslant L \\ k_i \geqslant M q_{\alpha_i}}} H\left(\frac{k_i}{M}, q_{\alpha_i}\right) > \frac{1}{M} \ln \frac{L N_{\mathcal{B}}}{\varepsilon}, \tag{3}$$

the pair (u,v) is ε-meaningful.

In this corollary, it appears clearly that the values of k such that (u,v) is ε-meaningful only depends on the logarithm of L, $N_{\mathcal{B}}$ and ε. In practice, we choose L about 32 which is compatible with our perceptual accuracy of directions. We also take $\varepsilon = 1$ since it means that we have in average less than 1-false detection. But, as we shall see, really similar images have much smaller NFA and the choice of ε is not really important. Thus, in all experiments, we always set $\varepsilon = 1$.

4 Random sampling

4.1 Problem statement

The a contrario model assumes that the values $D(x)$ are i.i.d. in $(0, \pi)$. This implicitly means that it is assumed that the direction $\theta_u(x)$ and $\theta_v(x)$ are independent for any x, and that these directions are independent when x describes the image plane. The NFA is nothing but a measure of the deviation to this hypothesis. If a few points are randomly drawn in the image, this assumption is clearly reasonable. However, since natural images contain alignments the second assumption becomes clearly false if we sample too many points. Moreover, if the two images have a casual alignment in common, this segment will induce a very strong deviation from the independence assumption, and can be detected. We then face the following dilemma for choosing the number of samples M:

- it must be large enough to allow us to contradict the independence hypothesis and to obtain small values of the number of false alarms for two similar images.
- it must be small enough to avoid the "common alignment problem".

In order to evaluate the typical magnitude of the number of sample points, let us assume that v differs from u by an additive Gaussian noise $\mathcal{N}(0, \sigma^2)$, which will be our hypothesis $\mathcal{H}_0$. We insist that we use this $\mathcal{H}_0$ to determine the number of sample points, but since we cannot assert that this model is realistic, the detection eventually relies only upon the background model $\mathcal{H}_1$. By computing the gradient by a finite difference scheme, it is possible to assume that the gradient coordinates of v are also corrupted by a white Gaussian noise (with a variance depending on the numerical scheme). If the law of the gradient norm is empirically estimated, it becomes possible to compute the law of the direction variation D, $P_{\mathcal{H}_0}(D < \alpha)$.

4.2 Bounds on the number of sample points

By definition, we detect the pair (u, v) as ε-meaningful, if $NFA(u, v) < \varepsilon$. Hence, we would like the value $P(NFA(u, v) < \varepsilon | \mathcal{H}_0)$ to be large whenever v is a noisy version of u. Assume also that u is an image of a query base $\mathcal{Q}$ containing $N_{\mathcal{Q}}$ images (and v is still in the database $\mathcal{B}$). If we want less than ε detection in the *a contrario* model by comparing all the pairs in $\mathcal{Q} \times \mathcal{B}$, we have to multiply the NFA definition (1) by $N_{\mathcal{Q}}$. Let

$$k_\alpha = \inf\{k, \text{ s.t. } N_{\mathcal{Q}} \cdot N_{\mathcal{B}} \cdot L \cdot B(M, k, q_\alpha) < \varepsilon\}.$$

To make things simpler, assume that we compute the NFA with only one value of angle α (so that $L = 1$). Since there is no ambiguity, we drop the subscript α. If K is the random

number of points such that $D < \alpha$, the pair (u, v) is detected if an only if $K \geqslant k$. The probability of detection under $\mathcal{H}_0$ is therefore

$$P_D \equiv P(K \geqslant k|\mathcal{H}_0) = B(M, k, p). \tag{4}$$

where

$$p = P_{\mathcal{H}_0}(D < \alpha),$$

which is known, since we have a model of noise.

Definition 2 *We call number of misses*

$$\mathcal{M}(M, k) = N_{\mathcal{Q}} N_{\mathcal{B}} (1 - B(M, k, p)). \tag{5}$$

As for the number of false alarms, if $\mathcal{M}(M, k) < \varepsilon$, it is clear that the expected number of misdetections under hypothesis $\mathcal{H}_0$ is less than ε.

The noise model clearly implies that p (the probability that gradient directions are alike when both images are the same) is larger than q (probability that the directions are alike for images of noise, i.e. the *a contrario* model) unless the images are constant of $\sigma = +\infty$, which is of little interest, and $p \to q$ when $\sigma \to +\infty$ (up to a normalization of grey level, the image tends to a white noise).

From estimates on the tail of the binomial law, we obtain the following necessary conditions on the number of samples M.

Proposition 3 *Assume that $\mathcal{M}(M, k) < \varepsilon$. Then, for some positive constant $C \simeq 0.39246$,*

$$M(p - q)^2 \geqslant \min(p(1 - p), q(1 - q)) \left(C + \ln \frac{N_{\mathcal{Q}} N_{\mathcal{B}}}{\varepsilon \sqrt{M}} \right). \tag{6}$$

Proof. From (4), we know that $1 - P_D = B(M, M - k, 1 - p)$. A refined Stirling inequality [3] implies that

$$\begin{aligned} \frac{\varepsilon}{N_{\mathcal{B}} N_{\mathcal{Q}}} &> B(M, M - k, 1 - p) \\ &\geqslant \binom{M}{M - k} (1 - p)^{M-k} p^k \\ &\geqslant \frac{2}{\sqrt{2\pi M}} e^{-1/6} e^{-MH(1-k/M, 1-p)}. \end{aligned}$$

Thus

$$M \cdot H\left(1 - \frac{k}{M}, 1 - p\right) > C + \ln \frac{N_{\mathcal{B}} N_{\mathcal{Q}}}{\varepsilon \sqrt{M}},$$

with $C = \frac{1}{6} + \frac{1}{2}\ln\frac{\pi}{2} \simeq 0.39246$. Since $k > Mq$, we also have $H\left(1 - \frac{k}{M}, 1-p\right) < H(1-q, 1-p)$. By convexity of H,

$$\begin{aligned} H(1-q,1-p) &\leqslant (p-q)\partial_x H(1-q,1-p) \\ &= (p-q)\ln\left(\frac{1-q}{q}\frac{p}{1-p}\right). \end{aligned}$$

Moreover

$$\begin{aligned} \ln\left(\frac{1-q}{q}\frac{p}{1-p}\right) &= \int_q^p \frac{dx}{x(1-x)} \\ &\leqslant (p-q)\max_{x\in[p,q]}\frac{1}{x(1-x)}. \end{aligned}$$

Since the function on the right hand side is convex, it attains its maximum on the boundary of the interval, and this completes the proof. □

The estimate above tells that, when the noise amount σ becomes large, M grows like $\frac{1}{(p-q)^2}$. This is not strictly exact because of the $\ln M$ terms on the right side of (6). This term is unavoidable since it appears in any sharp lower bound of the tail of the binomial law. In the following result, it will be proved that the order of magnitude $O((p-q)^{-2})$ is sufficient.

The first sufficient condition below gives an upper bound to k.

Lemma 1

$$k \leqslant 1 + Mq + \left(\frac{M}{2}\left(\ln\frac{LN_{\mathcal{B}}N_{\mathcal{Q}}}{\varepsilon}\right)\right)^{1/2}. \tag{7}$$

Proof. Since $k = \inf\{j \text{ s.t. } N_{\mathcal{B}}N_{\mathcal{Q}}\cdot B(M,k,q) < \varepsilon\}$, $B(M,k-1,q) > \frac{\varepsilon}{N_{\mathcal{B}}N_{\mathcal{Q}}}$ holds, also yielding

$$H\left(\frac{k-1}{M}, q\right) < \frac{1}{M}\ln\frac{N_{\mathcal{B}}N_{\mathcal{Q}}}{\varepsilon}.$$

Convexity properties of the entropy H yield $H(r,q) \geqslant 2(r-q)^2$. Setting $r = \frac{k-1}{M}$ gives the result. □

It is then possible to prove the following sufficient condition on the number of samples M.

Proposition 4 *If*

$$M \geqslant \frac{2}{(p-q)^2}\ln\frac{N_{\mathcal{B}}N_{\mathcal{Q}}}{\varepsilon}. \tag{8}$$

then $\mathcal{M}(M,k) < \varepsilon$.

Proof. If M is large enough, we can assume that $k < Mp$ from (7). A sufficient condition to $\mathcal{M}(M, P) < \varepsilon$ is

$$H\left(1 - \frac{k}{M}, 1 - p\right) > \frac{1}{M} \ln \frac{N_{\mathcal{B}} N_{\mathcal{Q}}}{\varepsilon}$$

Since by convexity $H(r, p) \geqslant 2(r - p)^2$, it suffices that

$$2\left(p - \frac{k}{M}\right)^2 \geqslant \frac{1}{M} \ln \frac{N_{\mathcal{B}} N_{\mathcal{Q}}}{\varepsilon},$$

which is implied by

$$p - q - \left(\frac{1}{2M} \ln \frac{N_{\mathcal{B}} N_{\mathcal{Q}}}{\varepsilon}\right)^{1/2} > \left(\frac{1}{2M} \ln \frac{N_{\mathcal{B}} N_{\mathcal{Q}}}{\varepsilon}\right)^{1/2},$$

and the result directly follows. □

5 Numerical applications and experiments

5.1 Justification of the background model

The background model should be sound for two unrelated images. Let us make the following experiment. Let us compute the empirical distribution of the gradient direction on two images. Because of quantization and presence of strongly privileged directions, these two histograms are not uniform at all. Nevertheless, the distribution of the difference of the direction, taken at *two* random locations (that is different points in the two images) is the circular convolution of these histograms. On many pairs of images, we indeed check that the difference of the repartition function with a uniform distribution in $(-\pi, \pi)$ is everywhere less than 0.01.

5.2 Number of sample points under hypothesis $\mathcal{H}_0$

On Fig 1, we discuss the relation between σ (the noise standard deviation), M (the number of sample points) and the detection rate as explained in Sect. 4.2. By varying σ and M, we empirically retrieve the bound estimate of Sect. 4.2.

5.3 Experiments of image retrieval

We first consider the following experiment. We select a single image in a sequence containing about one hour of video (86096 images). A white Gaussian noise with standard

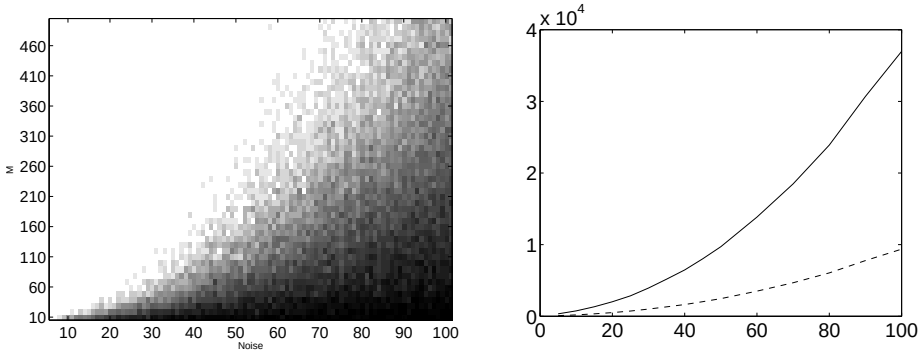

Figure 1: We match an image with some of its corrupted versions by a white Gaussian noise, for σ varying between 5 and 100 (horizontal axis), and for a number of samples M between 10 and 500 (vertical axis). For each couple (σ, M), 50 trials are drawn, yielding $N_{\mathcal{B}} = 250000$. The grey level in the left plot is the number of detections (white for 50 and black for 0). The curves on the right are the suffi cient and necessary values of M for controlling the number of misses, given by (6) and (8) respectively. As expected, the empirical results on the left are between these curves and bounds are not sharp.

deviation $\sigma = 30$ is added to this image, and will be taken as the query. The proposed criterion is applied with $N = 500$ random sample points in the images. The true image was detected with a NFA equal to 10^{-14}. About 20 images (belonging to the same static shot) are detected around the query. There was a single true false alarm (unrelated images) with a NFA equal to $10^{-0.73}$, which was probably due to the presence of the logo. No false alarms were obtained for an impulse noise of 50%. Extreme JPEG compression (quality less than 10) may lead to false detections since gradient orientation is constrained by the blocking effect. For usual compression ratio, this effect was not observed.

On Fig. 3, two images of a movie are compared. The scene exhibits a strong transparency effect and an important contrast change. Thus, the grey levels in those images are different. But this is not a good criterion at all, since the images clearly have a common cause. The direction comparison proves that these images are similar in the sense that there resemblance cannot be explained by the *a contrario* model. It was empirically checked that sample points were quite uniformly distributed in the images.

Fig. 4 shows the robustness to occlusion. The score panel occludes a large part of the image in this video of tennis match. The two images are detected as very similar since their number of false alarms is about 10^{-50}. Since an hour of video contains about 10^5 images, the match remains meaningful for any size of database. The threshold on the gradient norm is equal to 5 in this experiment. If we take it equal to 0.2 (still with 200 sample points),

Figure 2: The middle image is a 50% impulse noise version of the original one. In a database of 10^5 images, they still match with a NFA close to 10^{-5}. The right plot shows $-\log_{10}(NFA)$ for the first 50000 images of the sequence, the query being the noisy image. The peaks indeed correspond to exactly the same view of the stadium. The same views, but translated by 10 pixels, are not detected, since no prior registration has been performed.

Figure 3: Robustness to transparency. The two images are selected from a movie. The background is fixed, but the contrast changes a lot and a transparency layer is also moving. Nevertheless, with 200 sample points, $\log_{10}(NFA) = -43.2$, and images are thus detected as very similar.

the NFA increases since we select points where the gradient orientation is dominated by quantization. However, with an equal probability, we select points with larger gradients, and the directions then match very well. Therefore, the NFA is still very low, and about 10^{-32}.

Figure 4: Robustness to occlusion. Despite the large occlusion the two images are detected as very similar with $\log_{10}(NFA) = -50.1$. The right plot gives the position of the 200 sample points. There are not points in constant areas (because of the gradient threshold). However, some points are selected in the non-matching area (scores), but the NFA is still very low.

As a way to check the thresholds validity, let us apply exactly the same scheme to pairs of consecutive images in a video. We first register the images by using the robust multiresolution algorithm by Odobez and Bouthemy [12], (available on line) which computes a 2D parametric motion model that usually corresponds to the dominant image motion. The evolution of the NFA through time is represented on Fig. 5. As expected, similarity is important since NFA are always lower than 10^{-20}, except at very precise instants that correspond to shot changes.

6 Conclusion and perspectives

We describe a fast algorithm allowing to efficiently compare two images from a random sampling of points and used it for image retrieval in databases or in video stream. Actually, the argument is quite general and the thresholds are rigorously proved to be robust and can be fixed once for all, for any type of images. Hence the user does not have to tune any parameter. It could also be applied to parts of images instead of whole images, so that the methodology could be used in many applications of image retrieval, image matching or registration evaluation. These parts of images could be extracted from local characteristics as keypoints [11] or local frame based on stable directions [7, 13]. We could then estimate the same detection bounds for system similar to [14].

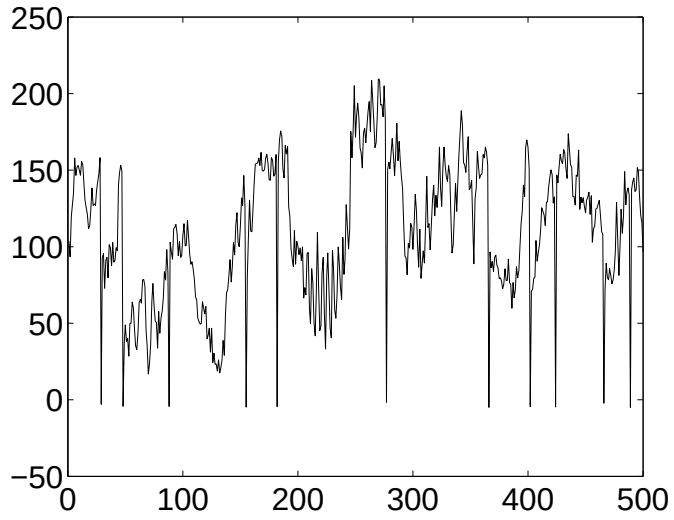

Figure 5: $-\log_{10}(NFA))$ between 500 consecutive pairs in a MPEG video sequence. Most of the time, the NFA is below 10^{-20}. The sudden drops correspond to shot changes. The NFA is a reliable value as predicted by Prop. 1.

References

[1] L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.

[2] A. Desolneux, L. Moisan, and J.M. Morel. A grouping principle and four applications. *IEEE Trans. on PAMI*, 25(4):508–513, 2003.

[3] W. Feller. *An Introduction to Probability Theory and its Applications*, volume I. J. Wiley, 3rd edition, 1968.

[4] M.A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24(6):381–395, 1981.

[5] W.E.L. Grimson and D.P. Huttenlocher. On the sensitivity of the Hough transform for object recognition. *IEEE Trans. on PAMI*, 12(3):255–274, 1990.

[6] W. Hoeffding. Probability inequalities for sum of bounded random variables. *J. of the Am. Stat. Assoc.*, 58:13–30, 1963.

[7] J.L. Lisani, L. Moisan, P. Monasse, and J.M. Morel. On the theory of planar shape. *SIAM Multiscale Mod. and Sim.*, 1(1):1–24, 2003.

[8] J.L. Lisani and J.M. Morel. Detection of major changes in satellite images. In *IEEE ICIP*, pages 941–944, 2003.

[9] D. Lowe. Object recognition from local scale-invariant features. In *IEEE ICCV*, pages 1150–1157, Corfu, 1999.

[10] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Comp. Vis.*, 60(2):91–110, 2004.

[11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. Submitted to International Journal of Computer Vision, 2005.

[12] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *J. of Vis. Comm. and Image Rep.*, 6(4):348–365, 1995.

[13] C.A. Rothwell. *Object Recognition Through Invariant Indexing*. Oxford Science Publications, 1995.

[14] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. In *Ninth IEEE ICCV*, pages 1470–1477, 2003.

[15] A. Venot, J.F. Lebruchec, and J.C. Roucayrol. A new class of similarity measures for robust image registration. *Comp. Vis. Graph. and Im. Proc.*, 28:176–184, 1982.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifi que,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399